

Exact Waiting Time Survival Function

Qamruz Zaman^{1,2*}, Alexander M. Strasak² and Karl P. Pfeiffer²

¹Department of Statistics, University of Peshawar, Pakistan

²Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Austria

Abstract

Although Kaplan-Meier survival function is the most commonly used statistical technique of survival analysis, it possesses a disadvantage. It may occur that Kaplan-Meier gives same survival probabilities for two groups having the same number of events and censored observations, although time spans between consecutive events (i.e. waiting times) may considerably vary. Therefore, severity of a disease, in terms of survival times, has no role in the conventional concept of Kaplan-Meier. To overcome this problem, in this paper we propose an exact waiting time survival function by explicitly considering waiting times between events. A new variance estimator, reducing to binomial variance in case of data free from censoring and time differences between two consecutive events equalling to 1, is presented. In order to compare the performance of the new estimator with conventional Kaplan-Meier estimator for small to large sample sizes, as well as for small to heavy censoring, we conducted a simulation study. The outcome shows that on average Pitman Closeness Criteria gives results in favour of our new estimator and confidence intervals have higher coverage rates, as compared to that obtained by Kaplan-Meier estimator, especially for lower confidence limits. Furthermore widths of confidence intervals are smaller than those based on Kaplan-Meier and Greenwood standard error. The proposed procedures are applied to a lung cancer data set.

Keywords: Kaplan-Meier survival function; Exact waiting time; Pitman closeness criteria; Confidence intervals; Greenwood variance estimator

Introduction

One of the basic purposes of survival analysis is the estimation of survival curves from censored survival data. In case of no covariates, non-parametric maximum likelihood estimator i.e. Kaplan-Meier (KM) survival function [1] is commonly used for estimating survival probabilities. However, the problem of Kaplan-Meier survival function is that it gives same survival probabilities for two groups having the same number of events and censored observations, by ignoring time spans (i.e. waiting times) between consecutive events.

To illustrate this, consider two groups X and Y each having n observations. In both groups, consider m events and n-m censored observations. Supposing the sequence of occurrence of events and censored observations to be equal, but their discrete observed times to be different, we consider that one data set consists of time ranges t_1, t_2, \dots, t_p and another of time ranges t_1, t_2, \dots, t_q respectively, where $q > p$. Applying Kaplan-Meier survival function in this setting, we obtain the same results for both groups, ignoring severity of disease. We can explain this idea in a more understandable form, from the following dummy example.

Suppose we have data sets of two groups (G_I, G_{II}) having different diseases. Both groups have the same sequence of occurrence of events and censored observations, except times of occurrence are different. We apply Kaplan-Meier survival function on both data sets (Table 1). Although observed times regarding events and censoring are clearly different in both groups, the probability columns give the same results, indicating a severe flaw of Kaplan-Meier survival function.

In order to overcome this problem, in this paper we present an exact waiting time survival function (EKM), based on discrete survival times, as well as a variance estimator based on exact waiting times. Although the concept of discrete survival times is not entirely new [2-5] until now

no attention has been given to different times of occurrence of events i.e. waiting times between two consecutive events. We conducted three different simulation studies to compare (1) the performance of Kaplan-Meier estimator and exact waiting time estimator by using the Pitman Closeness Criteria [6-8], (2) coverage rates of their lower confidence limits and (3) widths of their confidence intervals. We applied these methods to a lung cancer data set.

Methods

Kaplan-Meier estimator

Let X_1, X_2, \dots, X_n be the true survival times of n individuals having the distribution function $F(x) = P(X \leq x)$ and Y_1, Y_2, \dots, Y_n be the censoring times having the distribution function $G(y) = P(Y \leq y)$. Further, let the survival times X_i and censoring times Y_i assumed to be independent. Let $T_i = \min \{X_i, Y_i\}$ be the observed survival time and $\delta_i = I(X_i \leq Y_i)$ be the indicator function, whether the observed time is event or censored. The Kaplan-Meier product limit estimator is defined by

$$\hat{S}_{KM} = \prod \left(1 - \frac{r_i}{n_i} \right)^{\delta_i} \quad (1)$$

where n_i is the number of individuals who are alive just before time T_i , including those who are about to die at this time, and r_i denotes the number dying at this time. Kaplan and Meier used Greenwood [9] variance estimator for their survival function

***Corresponding author:** Qamruz Zaman, Department of Statistics, University of Peshawar, Pakistan, E-mail: ayanqamar@gmail.com

Received June 09, 2011; **Accepted** August 09, 2011; **Published** September 25, 2011

Citation: Zaman Q, Strasak AM, Pfeiffer KP (2011) Exact Waiting Time Survival Function. J Biomet Biostat 2:117. doi:[10.4172/2155-6180.1000117](https://doi.org/10.4172/2155-6180.1000117)

Copyright: © 2011 Zaman Q, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Serial number	GI				GII			
	Time	Events	Number at risk	Probabilities	Time	Events	Number at risk	Probabilities
1	2	0	7		4	0	7	
2	4	1	6	0.8333	16	1	6	0.8333
3	6	1	5	0.6667	64	1	5	0.6667
4	8	0	4		256	0	4	
5	10	0	3		1024	0	3	
6	12	1	2	0.3333	4096	1	2	0.3333
7	14	1	1	0.0000	16284	1	1	0.0000

Table 1: Fictive survival probabilities of two dummy groups.

Time	event	no. at risks	Waiting time	Product	Prob.	Survival Prob.
t_1	r_1	n_1	$\Delta_1=1$	$n_1\Delta_1$	$p_1 = \left(1 - \frac{r_1}{n_1\Delta_1}\right)$	$\hat{S}_{EKM}(1) = p_1$
t_2	r_2	n_2	Δ_2	$n_2\Delta_2$	$p_2 = \left(1 - \frac{r_2}{n_2\Delta_2}\right)$	$\hat{S}_{EKM}(2) = p_1 \times p_2$
\vdots	\vdots	\vdots	\vdots	\vdots		
t_n	r_n	n_n	Δ_n	$n_n\Delta_n$	$p_n = \left(1 - \frac{r_n}{n_n\Delta_n}\right)$	$\hat{S}_{EKM}(n) = p_1 \times p_2 \times \dots \times p_n$

Table 2: Layout of exact waiting time survival function.

$$Var\left(\hat{S}_{KM}\right)=\hat{S}_{KM}^2\sum\left\{\frac{r_i}{\left[n_i\left(n_i-r_i\right)\right]}\right\}\tag{2}$$

Just like Kaplan-Meier survival function, also Greenwood variance estimator does not consider the waiting time between events.

Exact waiting time survival function

Due to not paying any attention to waiting times between two consecutive events, we assume severity of diseases to be equal. If the waiting time of the occurrence of two events of disease A is always less than that of two events of disease B, disease A is assumed to be more severe than B and it's survival probability therefore should be less than the survival probability of disease B. However, when applying conventional Kaplan-Meier survival function we erroneously obtain same survival probabilities for both diseases. To overcome this problem we developed an exact waiting time survival function, assuming the following sequence of observed survival and censored times:

Starting time _____ Event _____ Event _____
End of the study

Let Δ_i denote the waiting time between two consecutive events with $i = 2,3,\dots,n$ and $\Delta_1 = 1$. We proceed as Δ_2 equalling to the waiting time between the first and second event and Δ_3 equalling to the time difference between the second and the third event. So on Δ_n denotes the time difference between the $(n-1)^{th}$ and n^{th} event if the data is free from censoring. If the data consists of censored observations the number of differences is less than the number of observed times. Δ depends

on the occurrence of events and Kaplan-Meier survival function gives the probabilities of these events by using the number of events and the number of persons at risk. As events are important parts of survival analysis and Δ is depending on events, Δ is also an essential part of survival analysis. Moreover Δ 's also effect survival probabilities, since with passage of time, the number of person's at risk and Δ (not always $\Delta=1$) varies. As Δ_i and n_i 's are related (large $\Delta_i (>1)$ and large n_i result in a higher survival probability), their product $n_i\Delta_i$ leads to corresponding results in terms of survival probabilities, but contains much more information about nature and severity of a disease. So, instead of taking n_i , we consider $n_i\Delta_i$ in our exact waiting time survival function and therefore the time difference between two events is directly related to the survival probabilities as well as to the nature of disease, meaning that a larger Δ results in a higher survival probability (indicating the less severity of disease) and vice versa. This concept is not incorporated in the conventional method of Kaplan-Meier survival function. However, if the difference between two consecutive events is 1, $n_i\Delta_i$ reduces to n_i and gives the same survival probability at time i , as that obtained by Kaplan-Meier survival function.

The basic layout of the above introduced procedure, considering discrete survival times, is shown in (Table 2).

By using the Product-Limit procedure the exact waiting time survival function is given by

$$\hat{S}_{EKM} = \prod_{i=1}^n \left[1 - \frac{r_i}{n_i\Delta_i}\right]\tag{3}$$

Just like Kaplan-Meier survival function, our exact waiting time function starts from 1 at time zero and it approaches the lower limit 0 if the last observed time is event and the last observed difference equals

	Cen. dist	%cen	n	Pitmann Closeness Criteria				
				d_1	d_3	d_5	d_7	d_9
E(4)	U(1,5)	0	15	0.996	1.000	1.000	1.000	1.000
			50	0.986	1.000	1.000	1.000	1.000
			100	0.995	1.000	1.000	1.000	1.000
E(1)	U(2, 25)	0.5	15	0.880	0.914	0.967	0.967	1.000
			50	0.718	0.810	0.856	0.992	1.000
			100	0.755	0.832	0.888	0.997	1.000
E(0.1)	U(2,25)	30	15	0.851	0.500	0.588	0.728	0.966
			50	0.529	0.655	0.824	0.907	0.978
			100	0.572	0.849	0.923	0.965	0.999
E(0.01)	U(2,25)	87	15	1.000	0.985	0.868	0.705	0.886
			50	1.000	0.969	0.829	0.515	0.463
			100	1.000	0.971	0.824	0.451	0.485
E(0.01)	U(2,5)	96	15	1.000	0.999	0.993	0.988	0.996
			50	1.000	1.000	0.999	0.994	0.927
			100	1.000	1.000	1.000	0.999	0.940
W(0.5, 1)	U(2, 25)	4	15	0.897	0.938	0.983	0.991	1.000
			50	0.861	0.825	0.912	0.997	1.000
			100	0.827	0.830	0.894	0.999	1.000
W(0.5, 50)	U(2, 25)	60	15	0.998	0.669	0.410	0.460	0.919
			50	0.950	0.472	0.401	0.552	0.795
			100	0.950	0.610	0.497	0.493	0.784
W(0.5, 100)	U(2, 25)	70	15	1.000	0.984	0.926	0.766	0.535
			50	1.000	0.915	0.787	0.587	0.548
			100	0.995	0.768	0.619	0.505	0.602
W(1.5, 5)	U(2, 25)	10	15	0.662	0.515	0.623	0.767	0.812
			50	0.809	0.802	0.844	0.912	0.995
			100	0.781	0.824	0.870	0.936	0.981
W(1.5, 25)	U(2, 25)	66	15	0.999	0.770	0.489	0.447	0.887
			50	0.982	0.588	0.399	0.525	0.694
			100	0.970	0.588	0.517	0.655	0.834
W(1.5, 50)	U(2,25)	86	15	1.000	0.978	0.842	0.673	0.873
			50	1.000	0.955	0.771	0.460	0.477
			100	1.000	0.956	0.770	0.411	0.520
Llog(1.5,2)	U(2,15)	2	15	0.822	0.958	0.991	0.991	1.000
			50	0.797	0.808	0.937	1.000	1.000
			100	0.748	0.792	0.947	1.000	1.000
Llog(1.5,2)	U(2,5)	5	15	0.869	0.958	0.990	0.989	1.000
			50	0.813	0.806	0.932	0.999	1.000
			100	0.769	0.785	0.930	1.000	1.000
Llog(1.5,1)	U(2,2.5)	20	15	0.849	0.847	0.896	0.886	0.985
			50	0.840	0.839	0.884	0.888	0.998
			100	0.796	0.780	0.886	0.910	1.000
Llog(0.5,1)	U(2,2.5)	40	15	0.954	0.827	0.847	0.968	0.993
			50	0.987	0.864	0.869	0.951	0.999
			100	0.999	0.810	0.841	0.957	1.000
Llog(1,1)	U(2, 50)	6	15	0.588	0.857	0.921	0.919	0.996
			50	0.663	0.826	0.882	0.953	1.000
			100	0.697	0.839	0.888	0.960	1.000
Llog(1,1)	U(2,2.5)	28	15	0.895	0.863	0.926	0.948	0.991
			50	0.943	0.825	0.896	0.931	0.999
			100	0.955	0.745	0.862	0.934	1.000

Table 3: Pitman Closeness Criterion for various distributions and percent censored at the selected deciles, based on 5000 simulated trials.

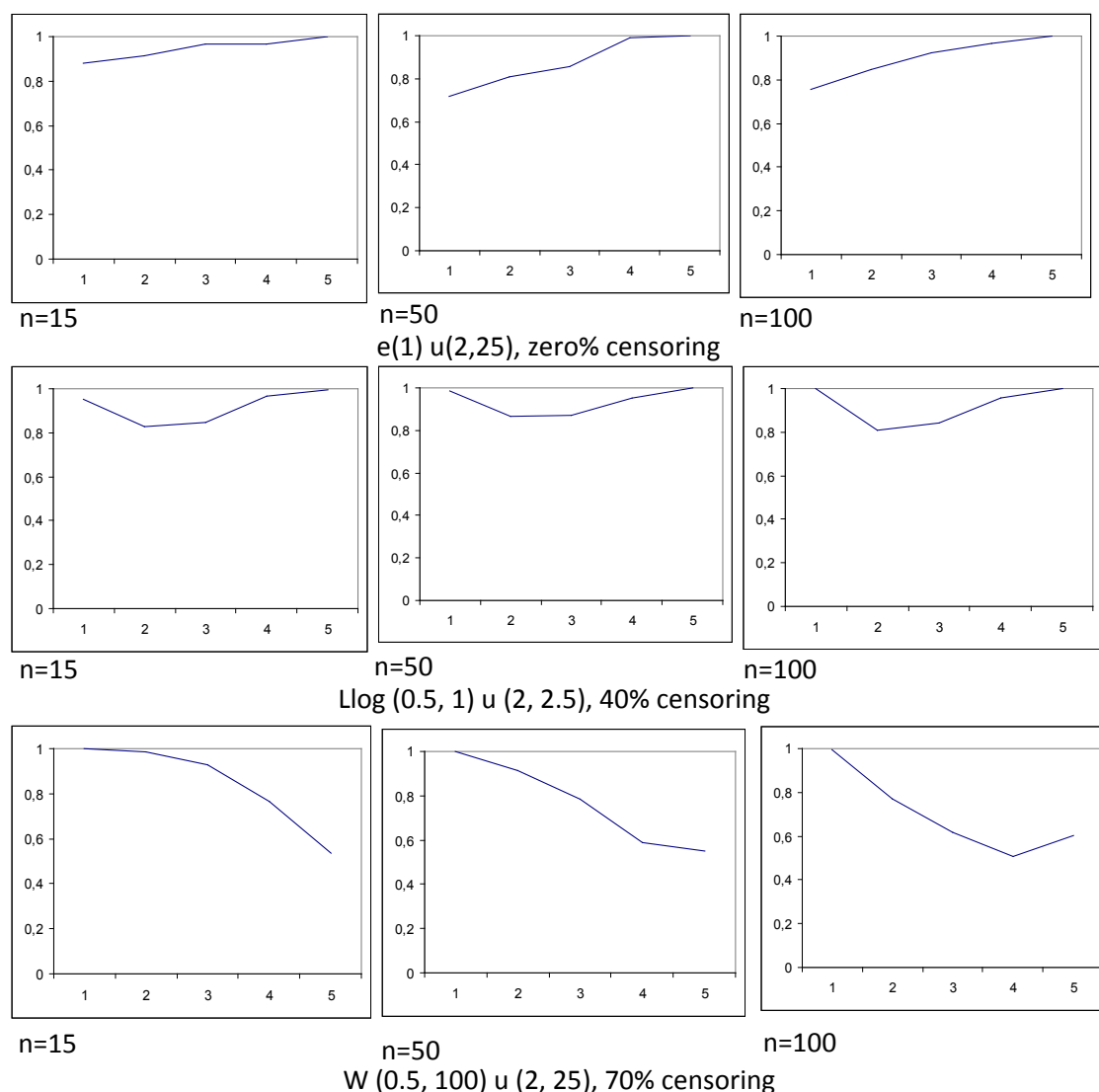


Figure 1: Pitman Closeness Criterion for various distributions and various percent censoring.

$\Delta = 1$. Otherwise, it is greater than 0. Unlike the conventional range of Survival function, the range of our proposed estimator is defined by

$$\left[\left(\frac{\Delta_n - 1}{\Delta_n} \right) \hat{S}_{EKM}(i-1), 1 \right]$$

If the n th difference $[\Delta_n]$ of our proposed estimator is 1, the range reduces to the range of Kaplan-Meier survival function. However, if this difference is greater than 1, the lower limit of our proposed estimator will always be greater than 0. In this case, our estimator approaches to the form of shrunk estimator. The idea of Shrunk has beneficial in improving the confidence interval coverage rate. The concept is not new in Survival analysis; for example, Parzen [10] proposed a ‘shifted piecewise linear’ empirical quantile function and Borkowf [11] proposed a Shrunk Survival estimator with the range $[(2n)^{-1}, 1 - (2n)^{-1}]$. Due to the behaviour of our new estimator, its lower confidence limits have higher coverage rates than Kaplan-Meier.

Variance of exact waiting time survival function

The variance estimator for our new procedure is obtained by using the delta method. By definition

$$\hat{S}_{EKM} = \prod_{i=1}^n \hat{p}_i \quad (4)$$

Taking log on both sides,

$$\log \hat{S}_{EKM} = \sum_{i=1}^n \log \hat{p}_i \quad \text{var} \left(\log \left(\hat{S}_{EKM} \right) \right) = \sum_i \text{var} \left(\log \hat{p}_i \right) \quad (5)$$

By applying delta method we derive

$$\text{var} \left(\log \left(\hat{p}_i \right) \right) \approx \left(\frac{1}{\hat{p}_i} \right) \frac{(1 - \hat{p}_i)}{n_i}$$

and so

Sur. dist	Cen. dist	%c	n	Lower coverage rate					
				d ₁		d ₂		d ₃	
				KM	EKM	KM	EKM	KM	EKM
E(4)	U(1,5)	0	15	0.701	0.701	0.231	0.231	0.237	0.237
			50	0.838	0.838	0.914	0.914	0.927	0.927
			100	0.818	0.818	0.990	0.990	0.999	0.999
E(0.1)	U(2,25)	30	15	0.831	0.968	0.321	0.753	0.430	0.580
			50	0.267	0.591	0.843	0.721	0.995	0.974
			100	0.532	0.515	0.998	0.987	1.000	1.000
E(0.01)	U(2,25)	87	15	1.000	1.000	0.998	1.000	0.983	0.999
			50	1.000	1.000	0.995	0.999	0.967	0.995
			100	1.000	1.000	0.996	0.999	0.966	0.992
E(0.01)	U(2,5)	96	15	1.000	1.000	1.000	1.000	1.000	1.000
			50	1.000	1.000	1.000	1.000	1.000	1.000
			100	1.000	1.000	1.000	1.000	1.000	1.000
W(0.5, 1)	U(2, 25)	4	15	0.196	0.411	0.017	0.054	0.553	0.562
			50	0.575	0.576	0.996	0.996	0.999	0.999
			100	0.995	0.996	1.000	1.000	1.000	1.000
W(0.5, 50)	U(2, 25)	60	15	0.998	0.999	0.865	0.990	0.643	0.953
			50	0.947	0.992	0.658	0.889	0.696	0.770
			100	0.883	0.955	0.644	0.741	0.860	0.865
W(0.5, 100)	U(2, 25)	70	15	1.000	1.000	0.957	0.998	0.809	0.986
			50	0.994	0.999	0.866	0.974	0.735	0.900
			100	0.979	0.992	0.807	0.906	0.773	0.821
W(1.5, 5)	U(2, 25)	10	15	0.388	0.544	0.071	0.133	0.593	0.553
			50	0.283	0.293	0.968	0.969	1.000	1.000
			100	0.930	0.932	1.000	1.000	1.000	1.000
W(1.5, 25)	U(2, 25)	66	15	0.999	1.000	0.933	0.995	0.758	0.966
			50	0.983	0.997	0.782	0.941	0.689	0.834
			100	0.953	0.982	0.705	0.844	0.783	0.806
W(1.5, 50)	U(2,25)	86	15	1.000	1.000	0.999	1.000	0.978	1.000
			50	1.000	1.000	0.994	1.000	0.955	0.993
			100	1.000	1.000	0.992	0.998	0.948	0.984
Llog(1.5,2)	U(2,5)	1.3	15	0.272	0.306	0.083	0.083	0.285	0.286
			50	0.503	0.512	0.944	0.944	0.967	0.967
			100	0.971	0.971	1.000	1.000	1.000	1.000
Llog(1.5,2)	U(2,15)	2.01	15	0.296	0.311	0.083	0.084	0.303	0.306
			50	0.456	0.460	0.943	0.943	0.973	0.973
			100	0.959	0.959	1.000	1.000	1.000	1.000
Llog(1.5,1)	U(2,15)	5	15	0.287	0.369	0.044	0.056	0.482	0.490
			50	0.425	0.455	0.978	0.978	0.999	0.999
			100	0.981	0.981	1.000	1.000	1.000	1.000
Llog(0.5,1)	U(2,25)	9	15	0.297	0.566	0.033	0.127	0.621	0.613
			50	0.375	0.399	0.994	0.993	1.000	1.000
			100	0.992	0.987	1.000	1.000	1.000	1.000
Llog(0.5,1)	U(2,2.5)	42	15	0.937	0.938	0.780	0.785	0.763	0.768
			50	0.998	0.999	0.999	0.999	0.999	0.999
			100	1.000	1.000	1.000	1.000	1.000	1.000
Llog(1,1)	U(2,25)	6	15	0.182	0.617	0.009	0.155	0.655	0.668
			50	0.625	0.568	0.999	0.998	1.000	1.000
			100	1.000	0.996	1.000	1.000	1.000	1.000
Llog(1,1)	U(2,2.5)	29	15	0.784	0.786	0.482	0.497	0.486	0.505
			50	0.812	0.812	0.811	0.811	0.815	0.815
			100	0.993	0.993	0.993	0.993	0.993	0.993

Table 4: Simulated asymptotic Normal 95 per cent lower bound coverage rates at the lower three deciles, with survival times generated from various distributions and censoring times from uniform distribution.

Sur. dist	Cen. dist	%c	Sum of Lower deciles widths						Sum of Upper deciles widths					
			d ₁		d ₂		d ₃		d ₇		d ₈		d ₉	
			KM	EKM	KM	EKM	KM	EKM	KM	EKM	KM	EKM	KM	EKM
E(4)	U(1,5)	0	363.6 274.2 139.5	363.6 274.2 139.5	1262.4 925.3 639.1	1262.4 925.2 639.0	1457.4 925.3 662.9	1457.4 925.3 662.9	1540.9 925.3 662.9	1540.9 925.2 662.9	1540.8 925.3 662.9	1540.8 925.2 662.9	1540.8 925.2 662.9	1540.8 925.2 662.9
E(0.1)	U(2,25)	30	268.6 783.0 753.0	67.7 547.4 722.4	1548.5 1341.8 974.4	698.6 1361.9 1003.4	2352.4 1485.3 1049.3	1697.8 1556.3 1056.2	2593.7 1376.4 967.1	2701.1 1380.1 967.2	2237.4 1220.3 859.9	2313.9 1222.0 859.9	1932.2 955.9 974.5	1878.6 956.3 674.6
E(0.01)	U(2,25)	87	0.0 0.0 0.0	0.0 0.0 0.0	0.7 6.4 10.6	0.0 0.6 2.3	72.7 112.5 131.8	2.8 16.3 36.4	1693.4 1825.8 1697.0	288.7 777.9 1063.8	2111.5 1909.1 1468.5	538.0 1191.8 1380.4	2180.9 1326.4 843.2	1010.5 1364.2 1035.2
W(0.5, 1)	U(2, 25)	4	1079.2 776.7 565.6	828.6 794.5 568.2	1835.7 1052.8 750.7	1820.6 1054.4 750.7	2097.1 1205.9 859.5	2123.6 1206.1 859.5	2464.7 1370.4 974.1	2464.6 1370.4 974.1	2440.7 1369.7 974.1	2440.7 1369.7 974.1	2437.2 1369.7 974.1	2437.2 1369.7 974.1
W(0.5, 25)	U(2, 25)	50	25.5 304.9 449.2	4.0 87.8 283.0	781.4 1232.6 1181.2	101.3 731.8 1075.4	1781.1 1765.2 1347.2	512.8 1578.4 1447.5	2831.5 1453.2 1013.6	2856.5 1489.4 1016.4	2389.7 1244.9 867.3	2557.6 1252.6 867.8	1994.8 939.2 659.5	1973.7 939.1 659.4
W(0.5, 50)	U(2, 25)	60	2.4 96.7 166.9	0.0 15.2 74.6	351.6 811.8 893.5	33.9 328.0 658.2	1210.1 1592.5 1509.7	209.7 95.4 1334.4	2999.2 1576.3 1083.8	2476.9 1766.8 1108.9	2597.1 1299.7 896.5	2578.2 1366.5 903.2	2085.6 952.2 662.9	2084.4 962.9 663.0
W(0.5, 100)	U(2, 25)	70	0.0 9.9 26.3	0.5 1.5 4.8	140.1 358.3 466.5	6.1 71.9 228.7	663.8 1116.9 1138.4	60.9 442.5 780.0	2852.5 1804.8 1211.3	1661.2 2064.8 1339.4	2731.5 1401.2 953.1	2129.2 1623.7 994.3	2219.1 987.6 675.4	2098.2 1044.4 683.2
W(0.5, 400)	U(2, 25)	84	0.0 0.0 0.0	0.0 0.0 0.0	8.9 33.3 38.9	0.0 2.5 8.5	138.4 245.1 286.7	2.0 51.0 110.7	2127.6 2039.2 1735.1	464.4 1239.6 1533.4	2432.4 1842.7 1273.3	870.2 1616.8 1468.0	2269.5 1190.7 755.8	1352.1 1415.5 882.4
W(1.5, 5)	U(2, 25)	10	892.4 781.2 578.8	709.1 794.3 580.1	1758.5 1076.1 771.1	1781.9 1079.4 771.2	2184.7 1246.2 886.8	2258.4 1246.9 886.8	2493.3 1351.6 956.9	2506.1 1351.9 956.9	2236.7 1239.7 881.6	2244.2 1239.9 881.6	1994.8 1010.6 700.6	1970.3 1010.2 700.6
W(1.5, 25)	U(2, 25)	66	1.4 29.3 67.5	0.4 6.6 30.1	208.5 545.4 684.2	27.3 181.5 402.8	871.4 1354.9 1405.5	147.2 695.6 1089.1	2979.6 1693.5 1147.0	2181.9 1925.8 1208.1	2686.6 1357.2 928.1	2531.5 1490.7 950.5	2165.9 975.7 672.6	2173.6 1015.2 680.1
W(1.5, 50)	U(2,25)	86	0.0 0.0 0.0	0.0 0.0 0.0	8.3 17.5 19.0	0.0 3.1 5.3	94.7 183.6 186.3	2.6 38.3 61.0	1873.7 1934.8 1770.8	396.3 982.4 1273.6	2260.9 1872.6 1388.7	736.3 1422.9 1469.6	2192.1 1264.6 794.1	1208.7 1407.1 942.9
Llog(1.5,2)	U(2,15)	1.9	921.8 704.1 511.1	933.1 706.2 511.2	1564.6 957.6 695.3	1566.4 957.6 692.2	1731.7 1030.1 738.2	1733.8 1030.1 738.2	2430.4 1371.0 974.6	2430.4 1371.0 974.6	2434.8 1371.1 974.7	2434.9 1371.0 974.6	2432.9 1371.0 974.6	2432.9 1371.0 974.6
Llog(1.5, 1)	U(2,2.5)	20	473.8 405.5 328.8	471.8 405.5 328.8	1066.2 813.9 626.2	1060.3 813.9 626.2	1487.8 885.0 632.6	1478.3 885.0 632.6	2464.4 1389.8 993.9	2467.0 1389.8 993.9	2321.4 1275.2 895.3	2322.1 1275.1 895.3	2157.5 1118.2 791.5	2157.6 1118.2 791.5
Llog(0.5,1)	U(2,2.5)	39	154.4 30.8 0.9	143.2 30.7 0.9	673.9 337.2 187.9	560.0 335.8 187.8	1315.7 956.3 771.3	1079.1 948.6 771.1	2561.6 1396.7 984.8	2525.9 1396.3 984.8	2364.0 1253.9 869.9	2348.1 1253.7 869.8	2222.9 1190.8 845.3	2216.2 1190.8 845.3
Llog(0.5,0.1)	U(2,2.5)	66.8	2.1 0.0 0.0	2.1 0.0 0.0	19.9 0.0 0.0	15.8 0.0 0.0	107.6 1.2 0.0	82.9 1.2 0.0	1997.8 1312.3 1001.0	1290.3 1115.7 964.4	2322.5 1455.7 1049.6	1690.0 1311.2 1025.4	2206.3 1176.7 854.9	1968.9 1136.8 848.0
Llog(0.5,0.01)	U(2,2.5)	86	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	298.9 17.3 0.7	108.8 10.7 0.6	818.9 246.7 102.9	306.2 116.6 68.0	1337.7 884.3 663.2	901.8 427.1 370.1
Llog(1,1)	U(2,50)	6	1146.8 809.4 589.6	564.2 795.6 602.4	1894.2 1083.4 773.6	1698.8 1092.2 774.2	2173.4 1235.6 878.1	2179.9 1236.6 878.2	2473.0 1344.3 946.0	2473.1 1344.3 946.0	2347.4 1283.6 906.0	2347.3 1283.6 906.0	2279.5 1273.1 904.8	2279.4 1273.1 904.8
Llog(1,0.1)	U(2,50)	32.9	216.6 784.3 780.1	6.9 117.9 478.8	1533.2 1395.1 1014.9	150.5 980.2 1163.2	2389.2 1533.2 1079.9	652.4 1697.0 1158.9	2609.5 1372.1 962.9	2769.2 1381.3 963.5	2234.6 1204.3 847.7	2369.6 1206.4 847.6	1914.1 936.9 660.6	1865.4 937.8 660.6

Table 5: Simulated table for the sum of three lower and three upper deciles, with data generated from various distributions.

$$\text{var} \left(\log \left(\hat{S}_{EKM} \right) \right) \approx \sum \left(\frac{1}{\hat{p}_i} \right) \frac{1 - \hat{p}_i}{n_i} \quad (6)$$

Now applying the delta method a second time on $\text{var} \left(\log \left(\hat{S}_{EKM} \right) \right)$ we get

$$\text{var} \left(\log \left(\hat{S}_{EKM} \right) \right) \approx \left(\frac{1}{\hat{S}_{EKM}} \right)^2 \text{var} \left(\hat{S}_{EKM} \right)$$

And by simplifying we derive our new variance estimator formula

$$\text{var} \left(\hat{S}_{EKM} \right) \approx \left[\hat{S}_{EKM} \right]^2 \sum_{i=1}^n \frac{r_i}{n_i (n_i \Delta_i - r_i)} \quad (7)$$

Unlike Greenwood's and Peto's [12] variance estimators, our new variance estimator uses the exact waiting time between two consecutive events. However, if the data is free from censoring and the differences of times between events equals 1, the new variance is reduced to binomial variance.

Simulation study

In order to compare the performance of our exact waiting time survival function with Kaplan-Meier survival function, we designed three different simulation studies. First, we compared Kaplan-Meier survival function with exact waiting time survival function in terms of Pitman Closeness Criterion; second, we compared confidence interval coverage rates; and third, the width of confidence intervals between both methods.

Comparison of exact waiting time survival function and kaplan-meier survival function by pitman closeness criterion

The Pitman Closeness Criterion [6] for two estimators at the selected points is defined as $PCC = P\{EKM - d_i \leq KM - d_i\}$

where d_i 's denote the deciles of the distribution at t . The criterion states that the absolute error of exact waiting time survival function is smaller than that of Kaplan-Meier survival function if $PCC > 0.5$. We used the uniform distribution $U(0, b)$ for generating censored survival times. As survival distributions, we choose (1) exponential, (2) weibull (with shape parameter 1.5, 0.5 i.e. monotone increasing and decreasing hazard rate) and (3) log-logistic (different shape parameter with monotone increasing, constant and decreasing hazard rate) distributions.

For comparison, we choose points of the form $F_x^{-1}(d_j)$ with fixed d_j 's ($d_1=0.1$, $d_3=0.03$, $d_5=0.5$, $d_7=0.7$, $d_9=0.9$) and generated 5000 data sets of survival times with various sample sizes ($n=15, 50, 100$). For each data set of survival times, the censoring times were generated from the uniform distribution. The observed survival times have been obtained from the censored and survival times. In order to get discrete times, we considered the zero decimal points. Pitman Closeness Criterion was calculated for each data set.

Table 3 and Figure 1 illustrate the results of our simulation study. At zero percent censoring, for all five deciles and sample sizes, the criterion gives results in favour of exact waiting time survival function. Also, by increasing censoring percentages, we observe similar results.

Only in very few points exact waiting time survival function gives more PCC-error than traditional Kaplan-Meier method. The same situation can be observed from (Figure 1) with selected sample sizes. These results hold for small as well as for large sample sizes and for small to heavy censoring percentages.

Lower confidence limit coverage rate

Considering the same censoring and survival distributions, we inspected the coverage properties of asymptotic normal confidence intervals. By using the estimates of the survival function and standard errors, we construct the asymptotic normal $(1-\alpha)$ 100 percent confidence intervals as

$$\text{estimated survival function} \pm \varphi^{-1} \left(1 - \alpha/2 \right) \text{estimated standard error}$$

We construct the confidence intervals by using Kaplan-Meier survival function and the Greenwood standard error. Similarly, we construct the confidence intervals centred on exact waiting time survival function and our proposed standard error. Since the upper limit of exact waiting time survival function is 1, the lower limit may not approach to the conventional limit i.e. 0. Moreover, as the upper coverage rates give the same results as that obtained by Kaplan-Meier survival function and Greenwood standard error, we are mainly interested in the comparison of the following two coverage rates.

$$\hat{S}_{KM} - \varphi^{-1} \left(1 - \alpha/2 \right) SE_G \left\{ \hat{S}_{KM} \right\} \quad (8)$$

and

$$\hat{S}_{EKM} - \varphi^{-1} \left(1 - \alpha/2 \right) SE \left\{ \hat{S}_{EKM} \right\} \quad (9)$$

We used the same censoring and survival distributions, but with different percentages of censoring. Table 4 shows the simulated 95% coverage rates of the lower three deciles, as there are more chances of values falling below or equalling to zero. At zero censoring, both methods give the same coverage rate. This may be induced by the fact that the waiting times between events are close to 1. However, with an increasing percentage of censoring, the new method gives better coverage. Only in very few cases and for small censoring percentages, at some points, its coverage rates are smaller than that obtained by the traditional methods

Width of confidence limits

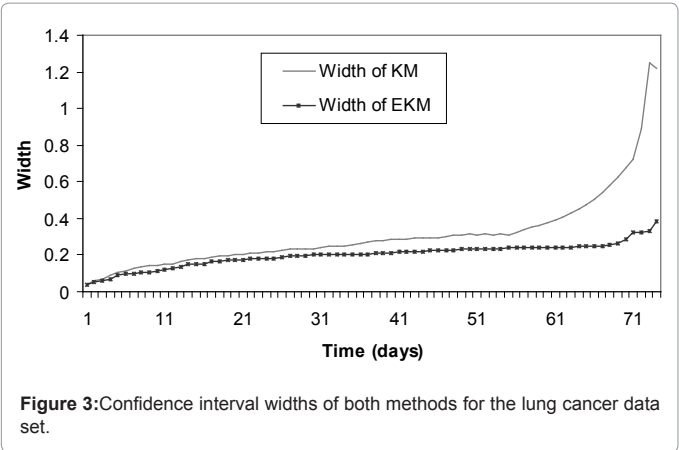
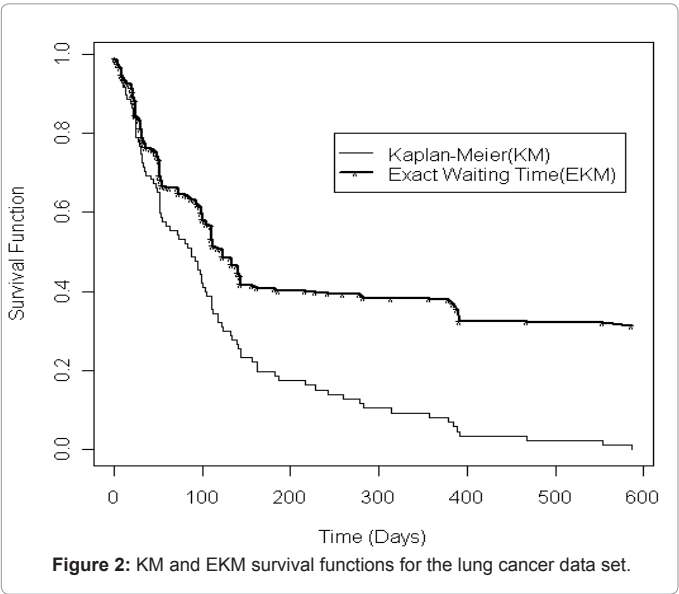
A further simulation study was performed in order to examine the width of the intervals for the same censoring and survival distributions. We calculated the width for both methods by using the simple logic "Upper limit - lower limit" for each simulation and took the sum over all simulations. The lower the width, the better the estimators. We considered the lower three deciles (d_1, d_2, d_3) and the upper three deciles (d_7, d_8, d_9), as in case of very heavy censoring, the lower deciles and their standard errors in most cases are undefined. Therefore, we also included the upper deciles to reach a general decision of the overall performance of both methods. The results of our simulation are

summarized in (Table 5). In censored free data, both methods gave the same results, but once again, as censoring increases, our new estimators give much smaller width than Kaplan-Meier and Greenwood standard error.

Application to a lung cancer data

We compared Kaplan-Meier survival function and exact waiting time survival function on a data set from the Veterans Administration lung cancer trial (presented by Prentice [13] and also used by Gupta [14]), where chemotherapy was induced to males with advanced inoperable lung cancer. For convenience, we considered the same part of the data set that Gupta used in his paper, consisting of 97 patients out of which 91 were events and 6 were censored. The survival times, given in days, were as following:

72	228	10	110	314	100*	42	144	30	384	4	13	123*
97*	59	117	151	22	18	139	20	31	52	18	51	122
27	54	7	63	392	92	35	117	132	162	3	95	162
216	553	278	260	156	182*	143	105	103	112	87*	242	111
587	389	33	25	357	467	1	30	283	25	21	13	87



7	24	99	8	99	61	25	95	80	29	24	83*	31
51	52	73	8	36	48	7	140	186	19	45	80	52
53	15	133	111	378	49							

* denotes a censored observation.

Table 6 (included as supplementary data) summarizes the data and methods in 11 columns, where column 1 shows the time in days, column 2 the number of person’s at risk (n_i), followed by the event column 3 (r_i). Column 4 shows the number censored (c_i) at different times. Δ is the heading of column 5 and represents the waiting time between two consecutive events. As the first Δ is 1, the difference between first and second waiting time is $3-1=2$ days, and so on we calculated the other waiting times. The next two columns represent Kaplan-Meier survival function and exact waiting time survival function, whereby both have the same value at first event, i.e. the first observed time. The important role of the concept of waiting times starts at the next stage, as from here both methods give different results in terms of survival probabilities This difference continues to exist through to the end of the analysis. Since the last observed time 587 (days) is an event, Kaplan-Meier survival function gives the value zero, while exact waiting time survival function yields a survival probability at that time equalling to 0.312. This is due to the fact that the waiting time between the last two observed events is greater than 1, i.e. $587-553=34$.

Lower confidence limits of Kaplan-Meier survival function and exact waiting time survival function are shown in columns 8 and 9. Here, we do not round the negative limit to zero, as we want to check the full behaviour of the two methods. We can see from the columns that the lower limit constructed by Kaplan-Meier gives 16 negative limits, while when constructed by exact waiting time survival function all these values are greater than zero. This shows that exact waiting time survival function is a type of left Shrunken Kaplan-Meier survival function and gives better results at this end. The next two columns give the upper limits of both methods and again we give the full estimates. In case of Kaplan-Meier survival function, there are 3 values greater than 1, while our exact waiting time survival function gives 4 values, due to a higher survival probability by this method at that point. Except this, the other limits settle down to routine. The last two columns give the widths of intervals for the two methods. We can see that at each time the width of confidence intervals constructed by exact waiting time survival function is smaller than that obtained by Kaplan-Meier survival function. These findings are also shown in (Figures 2 and 3).

Discussion

Kaplan-Meier survival function, which is a non-parametric technique of survival analysis, remains a reliable and frequently used method in medical research, as it is easy to understand, calculate and interpret. For the construction of confidence intervals, Greenwood variance estimator is commonly used. However, they both share the drawback of giving same results for two diseases of different nature, by ignoring waiting times between two consecutive events. In light of this deficiency, in this paper we proposed an exact waiting time survival function and a modified variance estimator. Unlike the traditional methods, these new estimators explicitly consider the waiting time between two events. The new methods perform equally well for

both, small and large sample sizes. Similarly, as censoring increases, the performance of our new estimators increases as well. The new estimators give better coverage rates for the lower confidence limits and yield in considerably smaller width of confidence intervals. Finally, the simplicity of these estimators make them attractive for use in different fields of medical research.

References

1. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53: 457-481.
2. Hogan JW, Laird NM (1998) Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Statistical Methods in Medical Research* 7: 28-48.
3. Brook Meyer R, Curriero FC (2002) Survival curve estimation with partial non-random exposure information. *Statistics in Medicine* 21: 2671-2683.
4. Lam FK, David IP (2003) REML and ML estimation for clustered grouped survival data. *Statistics in Medicine* 22: 2025-2034.
5. Adebayo SB, Fahrmeir L (2005) Analysing child mortality in Nigeria with geoaddditive discrete-time survival models. *Statistics in Medicine* 24: 709-728.
6. Keating JP, Mason RL, Sen PK (1993) Pitman's Measure of Closeness: A Comparison of Statistical Estimators. SIAM Philadelphia.
7. Rossa A, Zieliński R (1999) Locally Weibull-smoothed Kaplan-Meier estimator. *Institute of Mathematics Polish Academy of Sciences*.
8. Rossa A, Zieliński R (2002) A simple improvement of the Kaplan-Meier estimator. *Communication in statistics: Theory and Methods* 13: 147-158.
9. Greenwood Major (1926) A report on the natural duration of cancer. In *Reports on public Health and Medical Subjects*
10. Parzen E (1979) Nonparametric statistical data modelling. *Journal of the American Statistical Association* 74: 105-121.
11. Borkowf CB (2005) A simple hybrid estimator for the Kaplan-Meier survival function. *Statistics in Medicine* 24: 827-851.
12. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, et al. (1977) Design and analysis of randomized clinical trials requiring prolonged observation of each patient II Analysis and examples. *British Journal of Cancer* 35: 1-39.
13. Prentice RL (1973) Exponential survivals with censoring and explanatory variables. *Biometrika* 60: 279-288.
14. Gupta, RC (1999) A Study of Log-Logistic Model in Survival Analysis. *Biometrical Journal* 41: 431-443.